Journal of Multivariate Analysis **75**, 163–183 (2000)
doi:10.1006/jmva.2000.1908, available online at http://www.idealibrary.com on IDEAL®

Common Principal Components for Dependent Random Vectors

Beat E. Neuenschwander[1]

University of Berne, Switzerland

and

Bernard D. Flury[2]

Indiana University, Bloomington

Let the kp-variate random vector $\mathbf{X}$ be partitioned into k subvectors $\mathbf{X}_i$ of dimension p each, and let the covariance matrix $\mathbf{\Psi}$ of $\mathbf{X}$ be partitioned analogously into submatrices $\mathbf{\Psi}_{ij}$. The common principal component (CPC) model for dependent random vectors assumes the existence of an orthogonal p by p matrix $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}^t \mathbf{\Psi}_{ij} \boldsymbol{\beta}$ is diagonal for all (i, j). After a formal definition of the model, normal theory maximum likelihood estimators are obtained. The asymptotic theory for the estimated orthogonal matrix is derived by a new technique of choosing proper subsets of functionally independent parameters. © 2000 Academic Press

AMS 1991 subject classifications: 62H25, 62F10, 62F12.

Key words and phrases: asymptotic distribution, eigenvalue, eigenvector, entropy, maximum likelihood estimation, multivariate normal distribution, patterned covariance matrices.

1. INTRODUCTION

Principal component analysis as originally developed by Pearson (1901), Hotelling (1933), and Anderson (1963) is a one sample method. It is based on a reparameterization of a single covariance matrix $\mathbf{\Psi}$, the new parameters being the eigenvectors $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p$ and the eigenvalues $\lambda_1, ..., \lambda_p$ of $\mathbf{\Psi}$.

In recent years principal components have gone through some new developments which can be divided into three main categories:

[1] To whom correspondence should be adressed at: Department of Social and Preventive Medicine, University of Berne, Switzerland.

[2] B. D. Flury was killed in a tragic accident on July 6, 1999.

0047-259X/00 $35.00

(1) generalizations to several groups (Flury, 1988; Krzanowski, 1979), (2) nonlinear techniques (Gnanadesikan and Wilk, 1969; Gnanadesikan, 1977; Donnell *et al.*, 1994; Hastie and Stuetzle, 1989; Tibshirani, 1992), and (3) the use of principal components in other areas like discriminant analysis (Flury and Schmid, 1992), cluster analysis (Tarpey *et al.*, 1995), and regression analysis (Gunst and Mason, 1977; Mansfield *et al.*, 1977). This paper explores yet a new direction by relating principal components to patterned covariance matrices which arise whenever an inherent structure among variables is present. Although the dimension of the parameter space is often considerably reduced by the imposed constraints, estimation of patterned covariance matrices can be a difficult problem.

The model suggested here relies on the common principal component (CPC) model of Flury (1988). The CPC model assumes that the covariance matrices $\mathbf{\Psi}_1, ..., \mathbf{\Psi}_k$ of k independent groups can be diagonalized simultaneously by a common orthogonal matrix $\boldsymbol{\beta}$, i.e., $\mathbf{\Psi}_i = \boldsymbol{\beta}\mathbf{\Lambda}_i\boldsymbol{\beta}^t, i = 1, ..., k$. The CPC model was motivated by biometrical applications, where a pattern of similar principal components but possibly different variances for different species can be observed quite frequently.

The current paper discusses a common principal component model for the situation where the assumption of independence between the k groups is violated. This situation arises for example in twin studies or in the context of repeated measurements, where p measurements on the same unit are taken at k different points in time.

We assume that the random vector $\mathbf{X}$ is partitioned into k parts of dimension p each,

$$\mathbf{X} = [\mathbf{X}_1^t, \mathbf{X}_2^t, ..., \mathbf{X}_k^t]^t,$$

where each $\mathbf{X}_i$ is p-variate. Assuming that the first two moments exist, the mean vector and covariance matrix of $\mathbf{X}$ are partitioned analogously,

$$\mathrm{E}[\mathbf{X}] = [\boldsymbol{\mu}_1^t, \boldsymbol{\mu}_2^t, ..., \boldsymbol{\mu}_k^t]^t,$$

and

$$\mathrm{Cov}[\mathbf{X}] = \mathbf{\Psi} = [\mathbf{\Psi}_{ij}]_{i,\, j=1, ..., k} = \begin{bmatrix} \mathbf{\Psi}_{11} & \cdots & \mathbf{\Psi}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{\Psi}_{k1} & \cdots & \mathbf{\Psi}_{kk} \end{bmatrix}.$$

DEFINITION 1.1. The partitioned kp–variate random vector $\mathbf{X}$ with covariance matrix

$$\mathbf{\Psi} = [\mathbf{\Psi}_{ij}]_{i,\, j=1, ..., k}$$

satisfies the common principal component (CPC) model for dependent random vectors if there exists an orthogonal p by p matrix $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p]$ such that

$$\boldsymbol{\beta}^t \boldsymbol{\Psi}_{ij} \boldsymbol{\beta} = \boldsymbol{\Lambda}_{ij} = \operatorname{diag}(\lambda_{ij,1}, ..., \lambda_{ij,p})$$

is diagonal for all pairs (i, j), $i, j = 1, ..., k$.

An application of this model was presented by Klingenberg *et al.* (1996) where morphometric data on 88 female water striders, available for $p = 4$ variables in $k = 6$ discrete growth stages, were analyzed. The CPC model was mainly motivated by the fact that the orthogonal matrices arising from the spectral decompositions of the diagonal blocks $\mathbf{S}_{ii}$ $(i = 1, ..., 6)$ were close to each other. For a detailed analysis see Klingenberg *et al.* (1996). For further motivation see Flury and Neuenschwander (1995a).

However, it is not clear why one would want to model the off-diagonal blocks $\mathbf{S}_{ij}$ $(i \neq j)$ using the same orthogonal matrix. Besides the argument of parsimony, some motivation for this assumption arises from the principle of maximum entropy, introduced by Good (1963), which suggests entertaining the null hypothesis that has the maximum entropy (subject to some a priori constraints). We present this idea for $k = 2$ now: Assume that $\boldsymbol{\Psi}_{ii}$ have common principal components, i.e., $\Psi_{ii} = \boldsymbol{\beta} \boldsymbol{\Lambda}_{ii} \boldsymbol{\beta}^t$ $(i = 1, 2)$. Without loss of generality assume that $\boldsymbol{\Psi}_{11}$ and $\boldsymbol{\Psi}_{22}$ are both diagonal. The question then arises: under which choice of off-diagonal elements in $\boldsymbol{\Psi}_{12}$ does the random vector $\mathbf{X}$ have maximum entropy? Under normality assumptions, the answer is the CPC model for dependent random vectors. For $\mathbf{X} \sim N_{2p}(\boldsymbol{\mu}, \boldsymbol{\Psi})$, the entropy is $H(\mathbf{X}) = p \log(2\pi) + p + \log \det \boldsymbol{\Psi}/2$ (Kullback, 1959). By a straightforward generalization of Hadamard's inequality for positive definite matrices, we have

$$\det \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{bmatrix} \leqslant \det \begin{bmatrix} \boldsymbol{\Psi}_{11} & \operatorname{diag} \boldsymbol{\Psi}_{12} \\ \operatorname{diag} \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{bmatrix},$$

with equality exactly if $\boldsymbol{\Psi}_{12}$ is diagonal too. The proof for arbitrary k is analogous.

The results in this paper are derived under the assumption that $\mathbf{X}$ follows a multivariate normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$. Since the CPC model relates to the covariance matrix but not to the mean vector, we focus on $\boldsymbol{\Psi}$ only. Thus, starting with a sample of size N, we reduce it by sufficiency to the sample covariance matrix $\mathbf{S} = [\mathbf{S}_{ij}]_{i, j = 1, ..., k}$, which is distributed as Wishart with $n = N - 1$ degrees of freedom and scale matrix $\boldsymbol{\Psi}/n$. In Section 2 the maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_{ij}$ $(i, j = 1, ..., k)$ are derived, and Section 3 deals with their asymptotic distribution.

2. PARAMETER ESTIMATION

In the setup of Section 1, the log-likelihood function is given by

$$l(\mathbf{\Psi}) = C - \frac{n}{2}\,[\log\det\mathbf{\Psi} + \mathrm{tr}(\mathbf{\Psi}^{-1}\mathbf{S})],$$

where the constant C does not depend on $\mathbf{\Psi}$. We are going to present two versions of the log-likelihood function, which will simplify further calculations considerably. First note that the CPC model for dependent random vectors can be written as $\mathbf{\Psi} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^t$, where $\mathbf{B} := \mathbf{I}_k \otimes \boldsymbol{\beta}$ and $\mathbf{\Lambda} := [\mathbf{\Lambda}_{ij}]_{i,\,j=1,\,\ldots,\,k}$. Using the parameterization in $\boldsymbol{\beta}$ and $\mathbf{\Lambda}$, the log-likelihood funtion can be written as

$$l(\boldsymbol{\beta}, \mathbf{\Lambda}) = C - \frac{n}{2}\,[\log\det\mathbf{\Lambda} + \mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{B}^t\mathbf{S}\mathbf{B})].$$

We further define $\mathbf{\Lambda}_h^* := (\lambda_{ij,\,h})_{i,\,j=1,\,\ldots,\,k}\,(h = 1, \ldots, p)$ and $\mathbf{\Lambda}^* := \mathrm{bdiag}\,[\mathbf{\Lambda}_h^*]_{h=1,\,\ldots,\,p}$. Here we use bdiag to denote the block-diagonal operator

$$\mathrm{bdiag}\,[\mathbf{\Lambda}_h^*]_{h=1,\,\ldots,\,p} = \begin{bmatrix} \mathbf{\Lambda}_1^* & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2^* & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Lambda}_p^* \end{bmatrix}$$

Then

$$\mathbf{\Lambda}^* = \mathbf{I}_{(p,\,k)}\mathbf{\Lambda}\mathbf{I}_{(p,\,k)}^t = \mathbf{I}_{(p,\,k)}\mathbf{\Lambda}\mathbf{I}_{(k,\,p)},$$

where $\mathbf{I}_{(p,\,k)}$ is a commutation matrix (Henderson and Searle, 1979). By the diagonality of $\mathbf{\Lambda}^*$ we have $\mathbf{\Lambda}^{*-1} = \mathrm{bdiag}\,[\mathbf{\Lambda}_h^{*-1}]_{h=1,\,\ldots,\,p}$, but also, by the orthogonality of the commutation matrix, $\mathbf{\Lambda}^{*-1} = \mathbf{I}_{(p,\,k)}\mathbf{\Lambda}^{-1}\mathbf{I}_{(k,\,p)}$. Using the notation $\mathbf{\Lambda}^{-1} = [\mathbf{\Lambda}^{ij}]_{i,\,j=1,\,\ldots,\,k}$ for the inverse of $\mathbf{\Lambda}$, we see that all $\mathbf{\Lambda}^{ij}$ are diagonal, $\mathbf{\Lambda}^{ij} = \mathrm{diag}(\lambda^{ij,\,1}, \ldots, \lambda^{ij,\,p})(i, j = 1, \ldots, k)$, and it follows that $\lambda^{ij,\,h} = (\mathbf{\Lambda}_h^{*-1})_{ij}$ $(i, j = 1, \ldots, k; h = 1, \ldots, p)$. For the terms in the log-likelihood function related to the determinant and the trace, we obtain

$$\det\mathbf{\Lambda} = \det\mathbf{\Lambda}^* = \prod_{h=1}^{p} \det\mathbf{\Lambda}_h^*$$

and

$$\begin{aligned}\operatorname{tr}[\mathbf{\Lambda}^{-1}\mathbf{B}^t\mathbf{S}\mathbf{B}] &= \operatorname{tr}[\mathbf{\Lambda}^{*\,-1}\mathbf{I}_{(p,\,k)}\mathbf{B}^t\mathbf{S}\mathbf{B}\mathbf{I}_{(k,\,p)}] \\ &= \operatorname{tr}[\mathbf{\Lambda}^{*\,-1}\mathbf{Z}] = \sum_{h=1}^{p} \operatorname{tr}[\mathbf{\Lambda}_h^{*\,-1}\,\mathbf{Z}_{hh}],\end{aligned}$$

where

$$\mathbf{Z} := [\mathbf{Z}_{hg}]_{h,\,g=1,\,\dots,\,p} = \mathbf{I}_{(p,\,k)}\mathbf{B}^t\mathbf{S}\mathbf{B}\mathbf{I}_{(k,\,p)}$$

and $\mathbf{Z}_{hg} := (\boldsymbol{\beta}_h^t\,\mathbf{S}_{ij}\boldsymbol{\beta}_g)_{i,\,j=1,\,\dots,\,k}(h,\ g=1,\,\dots,\,p)$. Finally,

$$\operatorname{tr}(\mathbf{\Lambda}^{-1}\mathbf{B}^t\mathbf{S}\mathbf{B}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\operatorname{tr}(\mathbf{\Lambda}^{ji}\boldsymbol{\beta}^t\mathbf{S}_{ij}\boldsymbol{\beta}) = \sum_{h=1}^{p}\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda^{ij,\,h}\boldsymbol{\beta}_h^t\,\mathbf{S}_{ij}\boldsymbol{\beta}_h.$$

We summarize the results in the following lemma.

LEMMA 2.1. *In the CPC model for dependent random vectors, the following two versions of the likelihood function are equivalent*:

$$\begin{aligned}\text{(a)}\qquad l(\mathbf{\Lambda},\boldsymbol{\beta}) &= C - \frac{n}{2}\,[\log\det\mathbf{\Lambda} + \operatorname{tr}(\mathbf{\Lambda}^{-1}\mathbf{B}^t\mathbf{S}\mathbf{B})] \\ &= C - \frac{n}{2}\left[\log\det\mathbf{\Lambda} + \sum_{i=1}^{k}\sum_{j=1}^{k}\operatorname{tr}(\mathbf{\Lambda}^{ji}\boldsymbol{\beta}^t\mathbf{S}_{ij}\boldsymbol{\beta})\right],\end{aligned}$$

and

$$\begin{aligned}\text{(b)}\quad l^*(\mathbf{\Lambda}^*,\boldsymbol{\beta}) &= C - \frac{n}{2}\,[\log\ \det\ \mathbf{\Lambda}^* + \operatorname{tr}(\mathbf{\Lambda}^{*\,-1}\mathbf{Z})] \\ &= C - \frac{n}{2}\left[\sum_{h=1}^{p}\log\ \det\ \mathbf{\Lambda}_h^{*\,-1} + \sum_{h=1}^{p}\operatorname{tr}(\mathbf{\Lambda}_h^{*\,-1}\,\mathbf{Z}_{hh})\right],\end{aligned}$$

where $\mathbf{\Lambda}^* = \mathbf{I}_{(p,\,k)}\mathbf{\Lambda}\mathbf{I}_{(k,\,p)}$ *and* $\mathbf{Z} = \mathbf{I}_{(p,\,k)}\mathbf{B}^t\mathbf{S}\mathbf{B}\mathbf{I}_{(k,\,p)}$.

We refer to (a) and (b) as the *standard* and the *dual log-likelihood functions*, respectively.

When deriving an equation system for the maximum likelihood estimators of $\mathbf{\Lambda}$ and $\boldsymbol{\beta}$, we will tacitly assume that all eigenvectors $\boldsymbol{\beta}_h$ $(h=1,\,\dots,\,p)$ are well defined, i.e., we exclude the case of simultaneous sphericity. For the derivatives with respect to the elements of $\mathbf{\Lambda}$ we use the dual log-likelihood function and Lemma A.4. We obtain

$$\frac{\partial l^*(\mathbf{\Lambda}^*,\boldsymbol{\beta})}{\partial\mathbf{\Lambda}_h^{*\,-1}} = \mathbf{0} \Leftrightarrow 2\mathbf{\Lambda}_h^* - \operatorname{diag}\mathbf{\Lambda}_h^* - 2\mathbf{Z}_{hh} + \operatorname{diag}\mathbf{Z}_{hh} = \mathbf{0}\quad h=1,\,\dots,\,p,$$

which implies $\lambda_{ij,h} = \boldsymbol{\beta}_h^t \mathbf{S}_{ij} \boldsymbol{\beta}_h (i, j = 1, ..., k; h = 1, ..., p)$. For the derivatives with respect to the orthogonal matrix $\boldsymbol{\beta}$ we need a lemma whose proof is given in Appendix C.

LEMMA 2.2. *Let $\mathbf{C}_j$ be diagonal matrices, $\mathbf{C}_j = \operatorname{diag}(c_{j1}, ..., c_{jp})$, and let $\mathbf{A}_j$ be p by p matrices, $j = 1, ..., J$. Moreover, let $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p]$ be an orthogonal p by p matrix. Then the stationary points of the function*

$$f(\boldsymbol{\beta}) = \sum_{j=1}^{J} \operatorname{tr}[\mathbf{C}_j \boldsymbol{\beta}^t \mathbf{A}_j \boldsymbol{\beta}]$$

are the solutions to

$$\boldsymbol{\beta}_m^t \left[\sum_{j=1}^{J} (c_{jl} - c_{jm})(\mathbf{A}_j + \mathbf{A}_j^t) \right] \boldsymbol{\beta}_l = 0$$

for all $l, m = 1, ..., p$ $(l \neq m)$.

Applying Lemma 2.2 to the log–likelihood function leads to the following theorem.

THEOREM 2.3. *In the CPC model for dependent random vectors, the maximum likelihood estimators $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\beta}}$ are solutions of the equation system*

$$\boldsymbol{\beta}_m^t \left[\sum_{i=1}^{k} \sum_{j=1}^{k} (\lambda^{ij,l} - \lambda^{ij,m})(\mathbf{S}_{ij} + \mathbf{S}_{ji}) \right] \boldsymbol{\beta}_l = 0$$

and

$$\lambda_{ij,h} = \boldsymbol{\beta}_h^t \mathbf{S}_{ij} \boldsymbol{\beta}_h$$

for $l, m = 1, ..., p$ $(l \neq m)$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p]$ orthogonal, where the $\lambda^{ij,h}$ are the elements of $\boldsymbol{\Lambda}^{ij}$ in $\boldsymbol{\Lambda}^{-1} = [\boldsymbol{\Lambda}^{ij}]_{i,j=1,...,k}$ $(i, j = 1, ..., k, h = 1, ..., p)$.

For the numerical computation of the maximum likelihood estimators an algorithm developed by Flury and Neuenschwander (1995b) is available.

3. ASYMPTOTIC THEORY

In this section we derive the asymptotic distribution of the maximum likelihood estimators of Section 2. We use the well known result that under suitable regularity conditions the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $n\mathcal{I}_n^{-1}(\boldsymbol{\theta})$

(Serfling, 1980). Here $\hat{\boldsymbol{\theta}}_n$ denotes the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}$ and $\mathscr{I}_n(\boldsymbol{\theta})$ is the information matrix.

The derivation of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ differs from the classical approaches used by Anderson (1963) and Flury (1988). We only use $p(p-1)/2$ functionally independent elements of $\boldsymbol{\beta}$, denoted by the vector $\boldsymbol{\beta}^*$. This is the largest number of elements in $\boldsymbol{\beta}$ such that the asymptotic covariance matrix of the corresponding maximum likelihood estimators remains nonsingular. It should be noted, however, that $\boldsymbol{\beta}^*$ does not determine $\boldsymbol{\beta}$ completely, but rather up to finitely many $\boldsymbol{\beta}$'s only. For instance, for $p=2$, $\boldsymbol{\beta}^*$ consists of one element, but there are still four possible $\boldsymbol{\beta}$'s left. For $p=3$, there are eight possible versions of $\boldsymbol{\beta}$ for a given $\boldsymbol{\beta}^*$ (consisting of three elements). Furthermore, we make use of the dual parameter matrix $\boldsymbol{\Lambda}^*$ introduced in Section 2. Thus, the parameter vector $\boldsymbol{\theta}$ is given by

$$0=\begin{bmatrix} \operatorname{vech} \boldsymbol{\Lambda}_1^* \\ \vdots \\ \operatorname{vech} \boldsymbol{\Lambda}_p^* \\ \boldsymbol{\beta}^* \end{bmatrix}.$$

where "vech" is the vec-half operator (see Appendix A). The information matrix is partitioned as

$$\mathscr{I}_n(\boldsymbol{\theta})=\begin{bmatrix} \mathscr{I}_n(\operatorname{vech} \boldsymbol{\Lambda}_1^*) & \cdots & \mathscr{I}_n(\operatorname{vech} \boldsymbol{\Lambda}_1^*, \operatorname{vech} \boldsymbol{\Lambda}_p^*) & \mathscr{I}_n(\operatorname{vech} \boldsymbol{\Lambda}_1^*, \boldsymbol{\beta}^*) \\ \vdots & \ddots & \vdots & \vdots \\ \mathscr{I}_n(\operatorname{vech} \boldsymbol{\Lambda}_p^*, \operatorname{vech} \boldsymbol{\Lambda}_1^*) & \cdots & \mathscr{I}_n(\operatorname{vech} \boldsymbol{\Lambda}_p^*) & \mathscr{I}_n(\operatorname{vech} \boldsymbol{\Lambda}_p^*, \boldsymbol{\beta}^*) \\ \mathscr{I}_n(\boldsymbol{\beta}^*, \operatorname{vech} \boldsymbol{\Lambda}_1^*) & \cdots & \mathscr{I}_n(\boldsymbol{\beta}^*, \operatorname{vech} \boldsymbol{\Lambda}_p^*) & \mathscr{I}_n(\boldsymbol{\beta}^*) \end{bmatrix}$$

We shall make extensive use of vec- and vech-operators as well as of the commutation matrix $\mathbf{I}_{(p,\,p)}$, the duplication matrix $\mathbf{G}_p$, and the elimination matrix $\mathbf{H}_p$. Note that different versions of $\mathbf{H}_p$ are used by different authors; in our context we used the approach of Henderson and Searle (1979); see Appendix A. In fact, the main results of this section depend on the particular choice of $\mathbf{H}_p$. The following two lemmata will be helpful for further calculations (for their proofs see Appendix C).

LEMMA 3.1. *Let* $\mathbf{D}_p := \operatorname{diag}(\operatorname{vec} \mathbf{I}_p)$. *Then*

(i) $$\mathbf{H}_p(2\mathbf{I}_{p^2}-\mathbf{D}_p)=\mathbf{G}_p^t,$$

(ii) $$\mathbf{H}_p(\mathbf{I}_{p^2}+\mathbf{I}_{(p,\,p)}-\mathbf{D}_p)=\mathbf{G}_p^t,$$

(iii) $$\mathbf{G}_p^t(\mathbf{A}\otimes\mathbf{A})(\mathbf{I}_{p^2}+\mathbf{I}_{(p,\,p)})=2\mathbf{G}_p^t(\mathbf{A}\otimes\mathbf{A}),$$

(iv) $$\mathbf{H}_p(\mathbf{A}\otimes\mathbf{A})(\mathbf{I}_{p^2}+\mathbf{I}_{(p,\,p)})=2\mathbf{H}_p(\mathbf{A}\otimes\mathbf{A}).$$

LEMMA 3.2. *Let* $\mathbf{X}$ *be a symmetric p by p matrix with functionally independent variables (except* $x_{ij}=x_{ji}$*). Then*

(i) $$\frac{\partial \operatorname{vec} \mathbf{X}^{-1}}{\partial(\operatorname{vec} \mathbf{X})^t}=-(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1})(\mathbf{I}_{p^2}+\mathbf{I}_{(p,\,p)}-\mathbf{D}_p),$$

(ii) $$\frac{\partial \operatorname{vech} \mathbf{X}^{-1}}{\partial(\operatorname{vech} \mathbf{X})^t}=-\mathbf{H}_p(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1})\,\mathbf{G}_p.$$

We use Lemmata 2.1 and A.4 to obtain

$$\frac{\partial l^*(\boldsymbol{\Lambda}^*,\boldsymbol{\beta})}{\partial \boldsymbol{\Lambda}_h^{*\,-1}}=-\frac{n}{2}(-2\boldsymbol{\Lambda}_h^*+\operatorname{diag}\boldsymbol{\Lambda}_h^*+2\mathbf{Z}_{hh}-\operatorname{diag}\mathbf{Z}_{hh}).$$

Using $\operatorname{vec}(\operatorname{diag}\boldsymbol{\Lambda}_h^*)=\mathbf{D}_k\operatorname{vec}(\boldsymbol{\Lambda}_h^*)$, it follows that

$$\frac{\partial l^*(\boldsymbol{\Lambda}^*,\boldsymbol{\beta})}{\partial \operatorname{vec}\boldsymbol{\Lambda}_h^{*\,-1}}=\frac{n}{2}(2\mathbf{I}_{k^2}-\mathbf{D}_k)\operatorname{vec}\boldsymbol{\Lambda}_h^*+\mathbf{g}(\mathbf{Z}_{hh}),$$

where the vector $\mathbf{g}(\mathbf{Z}_{hh})$ does not depend on $\boldsymbol{\Lambda}$. Consequently,

$$\frac{\partial l^*(\boldsymbol{\Lambda}^*,\boldsymbol{\beta})}{\partial \operatorname{vech}\boldsymbol{\Lambda}_h^{*\,-1}}=\frac{n}{2}\mathbf{H}_k(2\mathbf{I}_{k^2}-\mathbf{D}_k)\operatorname{vec}\boldsymbol{\Lambda}_h^*+\mathbf{H}_k\mathbf{g}(\mathbf{Z}_{hh}).$$

Since the above expression does not depend on $\boldsymbol{\Lambda}_m^*$ $(m\neq h)$, it follows that

$$\mathscr{I}_n(\operatorname{vech}\boldsymbol{\Lambda}_h^{*\,-1},\operatorname{vech}\boldsymbol{\Lambda}_m^{*\,-1})=\mathbf{0}.$$

For the second derivatives, Lemma 3.2 implies

$$\frac{\partial^2 l^*(\boldsymbol{\Lambda}^*,\boldsymbol{\beta})}{\partial \operatorname{vech}\boldsymbol{\Lambda}_h^{*\,-1}\,\partial(\operatorname{vec}\boldsymbol{\Lambda}_h^{*\,-1})^t}=-\frac{n}{2}\mathbf{H}_k(2\mathbf{I}_{k^2}-\mathbf{D}_k)(\boldsymbol{\Lambda}_h^*\otimes\boldsymbol{\Lambda}_h^*)(\mathbf{I}_{k^2}+\mathbf{I}_{(k,\,k)}-\mathbf{D}_k).$$

Furthermore, by Lemma 3.1 we obtain

$$\begin{aligned}\frac{\partial^2 l^*(\boldsymbol{\Lambda}^*,\boldsymbol{\beta})}{\partial \operatorname{vech}\boldsymbol{\Lambda}_h^{*\,-1}\,\partial(\operatorname{vech}\boldsymbol{\Lambda}_h^{*\,-1})^t}&=-\frac{n}{2}\mathbf{H}_k(2\mathbf{I}_{k^2}-\mathbf{D}_k)\\&\quad(\boldsymbol{\Lambda}_h^*\otimes\boldsymbol{\Lambda}_h^*)(\mathbf{I}_{k^2}+\mathbf{I}_{(k,\,k)}-\mathbf{D}_k)\,\mathbf{H}_k^t\\&=-\frac{n}{2}\mathbf{G}_k^t(\boldsymbol{\Lambda}_h^*\otimes\boldsymbol{\Lambda}_h^*)\,\mathbf{G}_k.\end{aligned}$$

Hence $\mathscr{I}_n(\operatorname{vech}\boldsymbol{\Lambda}_h^{*\,-1})=n\mathbf{G}_k^t(\boldsymbol{\Lambda}_h^*\otimes\boldsymbol{\Lambda}_h^*)\,\mathbf{G}_k/2$. Using the relation between the information matrices of two related parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, namely $\mathscr{I}_n(\boldsymbol{\theta})=\mathbf{J}_{\phi,\,\theta}^t\,\mathscr{I}_n(\boldsymbol{\phi})\,\mathbf{J}_{\phi,\,\theta}$, where $\mathbf{J}$ denotes the matrix of partial derivatives,

$\mathbf{J}_{\phi,\theta} = \partial\phi/\partial\theta^t$, we obtain the information matrix of $\mathbf{\Lambda}_h^*$ by using Lemmata 3.2(ii), 3.1, and A.3(v),

$$\mathscr{I}_n(\text{vech}\,\mathbf{\Lambda}_h^*) = \frac{n}{2}\,\mathbf{G}_k^t(\mathbf{\Lambda}_h^{*\,-1}\otimes\mathbf{\Lambda}_h^{*\,-1})\,\mathbf{H}_k^t\,\mathbf{G}_k^t(\mathbf{\Lambda}_h^*\otimes\mathbf{\Lambda}_h^{*\,-1})$$
$$\times\,\mathbf{G}_k\mathbf{H}_k(\mathbf{\Lambda}_h^{*\,-1}\otimes\mathbf{\Lambda}_h^{*\,-1})\,\mathbf{G}_k$$
$$= \frac{n}{2}\,\mathbf{G}_k^t(\mathbf{\Lambda}_h^{*\,-1}\otimes\mathbf{\Lambda}_h^{*\,-1})\,\mathbf{G}_k.$$

Next we show that $\mathscr{I}_n(\text{vech}\,\mathbf{\Lambda}_h^*, \boldsymbol{\beta}^*)$ is zero ($h = 1, ..., p$), which will imply the asymptotic independence of $\hat{\mathbf{\Lambda}}_h^*$ and $\hat{\boldsymbol{\beta}}^*$. We have for an arbitrary function of $\boldsymbol{\beta}^*$, say $\mathbf{z} = \mathbf{z}(\boldsymbol{\beta}^*)$,

$$\frac{\partial}{\partial\mathbf{z}}\,\boldsymbol{\beta}_h^t\,\mathbf{S}_{ij}\boldsymbol{\beta}_h = \mathbf{J}_{\beta_h,z}^t\,\mathbf{S}_{ij}\boldsymbol{\beta}_h + \mathbf{J}_{\beta_h,z}^t\,\mathbf{S}_{ji}\boldsymbol{\beta}_h = \mathbf{J}_{\beta_h,z}^t(\mathbf{S}_{ij}+\mathbf{S}_{ji})\,\boldsymbol{\beta}_h.$$

Since $\mathbf{S}_{ij}$ is unbiased for $\mathbf{\Psi}_{ij}$, this implies (for $i \neq j$)

$$\mathrm{E}\left[\frac{\partial^2 l(\mathbf{\Lambda},\boldsymbol{\beta})}{\partial\lambda^{ij,h}\,\partial\mathbf{z}}\right] = -\frac{n}{2}\,\mathrm{E}\left[\frac{\partial}{\partial\mathbf{z}}\,\boldsymbol{\beta}_h^t(\mathbf{S}_{ij}+\mathbf{S}_{ji})\,\boldsymbol{\beta}_h\right]$$
$$= -\frac{n}{2}\,[\mathbf{J}_{\beta_h,z}^t(\mathbf{\Psi}_{ij}+\mathbf{\Psi}_{ji})\,\boldsymbol{\beta}_h] = -n\mathbf{J}_{\beta_h,z}^t\,\boldsymbol{\beta}\mathbf{\Lambda}_{ij}\,\boldsymbol{\beta}^t\boldsymbol{\beta}_h$$
$$= -n\mathbf{J}_{\beta_h,z}^t\,\boldsymbol{\beta}\mathbf{\Lambda}_{ij}\mathbf{e}_h = -n\lambda_{ij,h}\,\frac{\partial\boldsymbol{\beta}_h^t}{\partial\mathbf{z}}\,\boldsymbol{\beta}_h = \mathbf{0},$$

where $\mathbf{e}_h$ is the vector with 1 in position h, and zero elsewhere, and where we make use of

$$\mathbf{0} = \frac{\partial 1}{\partial\mathbf{z}} = \frac{\partial\boldsymbol{\beta}_h^t\,\boldsymbol{\beta}_h}{\partial\mathbf{z}} = 2\mathbf{J}_{\beta_h,z}^t\,\boldsymbol{\beta}_h = 2\,\frac{\partial\boldsymbol{\beta}_h^t}{\partial\mathbf{z}}\,\boldsymbol{\beta}_h.$$

A similar derivation holds for $i = j$. Thus

$$\mathscr{I}_n(\boldsymbol{\theta}) = \begin{bmatrix} \frac{n}{2}\,\mathbf{G}_k^t(\mathbf{\Lambda}_h^{*\,-1}\otimes\mathbf{\Lambda}_h^{*\,-1})\,\mathbf{G}_k & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \frac{n}{2}\,\mathbf{G}_k^t(\mathbf{\Lambda}_p^{*\,-1}\otimes\mathbf{\Lambda}_p^{*\,-1})\,\mathbf{G}_k & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathscr{I}_n(\boldsymbol{\beta}^*) \end{bmatrix}.$$

Finally, we obtain $\mathscr{I}_n^{-1}(\text{vech}\,\mathbf{\Lambda}_h^*) = 2\mathbf{H}_k(\mathbf{\Lambda}_h^*\otimes\mathbf{\Lambda}_h^*)\,\mathbf{H}_k^t/n$.

THEOREM 3.3. *In the CPC model for dependent random vectors, the p random vectors* $\sqrt{n}(\operatorname{vech} \hat{\boldsymbol{\Lambda}}_h^* - \operatorname{vech} \boldsymbol{\Lambda}_h^*)$, $h = 1, ..., p$, *are asymptotically distributed as* $N_{k(k+1)/2}(\mathbf{0}, \boldsymbol{\Sigma})$ *where*

$$\boldsymbol{\Sigma} = 2\mathbf{H}_k (\boldsymbol{\Lambda}_h^* \otimes \boldsymbol{\Lambda}_h^*)\, \mathbf{H}_k^t.$$

Moreover, they are independent of each other and independent of $\hat{\boldsymbol{\beta}}^*$.

Note that the definition of $\mathbf{H}_k$ as given in Appendix A is used in Theorem 3.3 rather than the definition of Magnus (1988).

For the derivation of the asymptotic distribution of $\hat{\boldsymbol{\beta}}^*$, we need the following lemma, whose proof is given in Appendix C.

LEMMA 3.4. *Let* $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p]$ *be an orthogonal p by p matrix, and* $u_1, ..., u_p$ *scalars, and let* $\mathbf{U}$ *be defined as follows*

$$\mathbf{U} := \operatorname{bdiag}[u_h \mathbf{I}_p]_{h=1, ..., p} = \operatorname{diag}(u_1, ..., u_p) \otimes \mathbf{I}_p.$$

Assume that β_{ml} *is an element of* $\boldsymbol{\beta}^*$, *and define*

$$\mathbf{b} := \operatorname{vec} \boldsymbol{\beta} \qquad \textit{and} \qquad \mathbf{c}_{ml} := \frac{\partial \operatorname{vec} \boldsymbol{\beta}}{\partial \beta_{ml}}.$$

Then

$$\frac{\partial \mathbf{c}_{ml}^t}{\partial \boldsymbol{\beta}^*} \mathbf{U}\mathbf{b} = -\frac{\partial \mathbf{b}^t}{\partial \boldsymbol{\beta}^*} \mathbf{U}\mathbf{c}_{ml}.$$

Using Lemmata 2.1 and A.1, we obtain $\operatorname{tr}(\boldsymbol{\Lambda}^{ij}\boldsymbol{\beta}^t\mathbf{S}_{ij}\boldsymbol{\beta}) = \mathbf{b}^t(\boldsymbol{\Lambda}^{ji} \otimes \mathbf{S}_{ij})\,\mathbf{b}$. Therefore the first derivative with respect to $\boldsymbol{\beta}^*$ is given by

$$\frac{\partial l(\boldsymbol{\Lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^*} = -\frac{n}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \mathbf{J}_{b, \beta^*}^t [(\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ij}) + (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ji})]\,\mathbf{b}.$$

The rows of $\mathbf{J}_{b, \beta^*}^t$ are $\mathbf{c}_{ml}^t = \partial \mathbf{b}^t / \partial \beta_{ml}$. Thus, for the second derivatives, we obtain

$$\frac{\partial \mathbf{c}_{ml}^t (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ij})\,\mathbf{b}}{\partial \boldsymbol{\beta}^{*t}} = \mathbf{b}^t (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ji})\, \mathbf{J}_{c_{ml}, \beta^*} + \mathbf{c}_{ml}^t (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ij})\, \mathbf{J}_{b, \beta^*}$$

and

$$\frac{\partial \mathbf{c}_{ml}^t (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ji})\,\mathbf{b}}{\partial \boldsymbol{\beta}^{*t}} = \mathbf{b}^t (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ij})\, \mathbf{J}_{c_{ml}, \beta^*} + \mathbf{c}_{ml}^t (\boldsymbol{\Lambda}^{ij} \otimes \mathbf{S}_{ji})\, \mathbf{J}_{b, \beta^*},$$

respectively. Taking the expectations and using Lemma A.1 yields

$$\begin{aligned} E\left[\frac{\partial \mathbf{c}_{ml}^t(\mathbf{\Lambda}^{ij}\otimes \mathbf{S}_{ij})\,\mathbf{b}}{\partial \boldsymbol{\beta}^{*t}}\right] &= \mathbf{b}^t(\mathbf{\Lambda}^{ij}\otimes \mathbf{\Psi}_{ji})\,\mathbf{J}_{c_{ml},\,\beta^*} + \mathbf{c}_{ml}^t(\mathbf{\Lambda}^{ij}\otimes \mathbf{\Psi}_{ij})\,\mathbf{J}_{b,\,\beta^*} \\ &= (\operatorname{vec}(\boldsymbol{\beta}\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}))^t\,\mathbf{J}_{c_{ml},\,\beta^*} + \mathbf{c}_{ml}^t(\mathbf{\Lambda}^{ij}\otimes \boldsymbol{\beta}\mathbf{\Lambda}_{ij}\boldsymbol{\beta}^t)\,\mathbf{J}_{b,\,\beta^*} \\ &= \mathbf{b}^t(\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}\otimes \mathbf{I}_p)\,\mathbf{J}_{c_{ml},\,\beta^*} + \mathbf{c}_{ml}^t(\mathbf{\Lambda}^{ij}\otimes \boldsymbol{\beta}\mathbf{\Lambda}_{ij}\boldsymbol{\beta}^t)\,\mathbf{J}_{b,\,\beta^*}, \end{aligned}$$

and we obtain the same result when $\mathbf{S}_{ij}$ is replaced by $\mathbf{S}_{ji}$. Since the matrix $\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}\otimes \mathbf{I}_p$ has the form of $\mathbf{U}$ in Lemma 3.4, we have

$$E\left[\frac{\partial \mathbf{c}_{ml}^t(\mathbf{\Lambda}^{ij}\otimes \mathbf{S}_{ij})\,\mathbf{b}}{\partial \boldsymbol{\beta}^{*t}}\right] = \mathbf{c}_{ml}^t(\mathbf{\Lambda}^{ij}\otimes \boldsymbol{\beta}\mathbf{\Lambda}_{ij}\boldsymbol{\beta}^t)\,\mathbf{J}_{b,\,\beta^*} - \mathbf{c}_{ml}^t(\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}\otimes \mathbf{I}_p)\,\mathbf{J}_{b,\,\beta^*}.$$

Putting all the rows together yields

$$E\left[\frac{\partial \mathbf{J}_{b,\,\beta^*}^t(\mathbf{\Lambda}^{ij}\otimes \mathbf{S}_{ij})\,\mathbf{b}}{\partial \boldsymbol{\beta}^{*t}}\right] = \mathbf{J}_{b,\,\beta^*}^t[(\mathbf{\Lambda}^{ij}\otimes \boldsymbol{\beta}\mathbf{\Lambda}_{ij}\boldsymbol{\beta}^t) - (\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}\otimes \mathbf{I}_p)]\,\mathbf{J}_{b,\,\beta^*}.$$

Hence the information matrix of $\boldsymbol{\beta}^*$ is given by

$$\mathscr{I}_n(\boldsymbol{\beta}^*) = n\sum_{i=1}^{k}\sum_{j=1}^{k}\mathbf{J}_{b,\,\beta^*}^t[(\mathbf{\Lambda}^{ij}\otimes \boldsymbol{\beta}\mathbf{\Lambda}_{ij}\boldsymbol{\beta}^t) - (\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}\otimes \mathbf{I}_p)]\,\mathbf{J}_{b,\,\beta^*}.$$

Since

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\mathbf{\Lambda}_{ij}\mathbf{\Lambda}^{ij}$$

is the sum of the diagonally placed blocks of $\mathbf{\Lambda}\mathbf{\Lambda}^{-1} = \mathbf{I}_{kp}$, we finally obtain

$$\begin{aligned} \mathscr{I}_n(\boldsymbol{\beta}^*) &= n\mathbf{J}_{b,\,\beta^*}^t\left[\sum_{i=1}^{k}\sum_{j=1}^{k}(\mathbf{\Lambda}^{ij}\otimes \boldsymbol{\beta}\mathbf{\Lambda}_{ij}\boldsymbol{\beta}^t) - (k\mathbf{I}_p\otimes \mathbf{I}_p)\right]\mathbf{J}_{b,\,\beta^*} \\ &= n\,\frac{\partial(\operatorname{vec}\boldsymbol{\beta})^t}{\partial \boldsymbol{\beta}^*}(\mathbf{I}_p\otimes \boldsymbol{\beta})\left[\sum_{i=1}^{k}\sum_{j=1}^{k}(\mathbf{\Lambda}^{ij}\otimes \mathbf{\Lambda}_{ij}) - k\mathbf{I}_{p^2}\right](\mathbf{I}_p\otimes \boldsymbol{\beta}^t)\,\frac{\partial \operatorname{vec}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{*t}}. \end{aligned}$$

THEOREM 3.5. *In the CPC model for dependent random vectors, the random vector* $\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}^*)$ *is asymptotically distributed as* $N_{p(p-1)/2}(\mathbf{0}, \mathbf{\Sigma})$, *where*

$$\mathbf{\Sigma}^{-1} = \frac{\partial(\operatorname{vec}\boldsymbol{\beta})^t}{\partial \boldsymbol{\beta}^*}(\mathbf{I}_p\otimes \boldsymbol{\beta})\left[\sum_{i=1}^{k}\sum_{j=1}^{k}(\mathbf{\Lambda}^{ij}\otimes \mathbf{\Lambda}_{ij}) - k\mathbf{I}_{p^2}\right](\mathbf{I}_p\otimes \boldsymbol{\beta}^t)\,\frac{\partial \operatorname{vec}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{*t}},$$

and $\boldsymbol{\beta}^*$ *consists of* $p(p-1)/2$ *functionally independent elements of* $\boldsymbol{\beta}$.

For a discussion of the term in $\boldsymbol{\Sigma}^{-1}$ relating to the matrix of derivatives we refer the reader to Appendix B. Theorems 3.3 and 3.5 may be applied in the usual way to the construction of confidence regions for parameters and Wald test statistics.

4. CONCLUSIONS

In this article we study a model of common principal components for patterned covariance matrices. Similarly to the ordinary CPC model, it provides a parsimonious parametrization of a "large" covariance matrix, which allows the investigator to reach simpler conclusions about the mechanisms of growth (Klingenberg *et al.*, 1996). Further simplification is achieved by imposing more constraints on the parameter space. For instance, if we assume that all submatrices $\boldsymbol{\Psi}_{ij}$ are proportional to $\boldsymbol{\Psi}_{11}$, we obtain the *proportional CPC model*

$$\boldsymbol{\Psi} = \mathbf{R} \otimes \boldsymbol{\Psi}_{11},$$

where $\mathbf{R}$ is a correlation matrix. In an even more restricted parametrization, $\mathbf{R}$ is assumed to have an equicorrelation structure, yielding the *equicorrelation CPC model*. Both of these models are studied in detail in Neuenschwander (1991) and illustrated on anthropometric data. On the other hand, one might want to relax the assumption of equality of all eigenvectors, i.e., assume that only some $q < p$ eigenvectors of the $\boldsymbol{\Psi}_{ij}$ are common. This leads to a partial CPC model similar to those studied in Flury (1988). To our knowledge, no results are available yet on partial CPC models in patterned covariance matrices. Another possible direction for future research is to study similar models for correlation matrices rather than covariance matrices.

In all of these models, the basic assumption is the existence of an orthogonal matrix that diagonalizes all $\boldsymbol{\Psi}_{ij}$ simultaneously. Much of the algebraic structure of the CPC model is preserved, however, if the assumption of orthogonality of $\boldsymbol{\beta}$ is replaced by nonsingularity, i.e., existence of a nonsingular matrix $\boldsymbol{\beta}$ such that all $\boldsymbol{\beta}'\boldsymbol{\Psi}_{ij}\boldsymbol{\beta}$ are diagonal. This leads to the model of *common canonical variates* studied by Neuenschwander and Flury (1995).

APPENDIX A: RESULTS FROM MATRIX ALGEBRA

We summarize here a few results used in Sections 2 and 3. Useful references are Henderson and Searle (1979) and Magnus (1988). The following three lemmata deal with vec- and vech-operators, commutation, elimination, and duplication matrices. For an arbitrary p by q matrix $\mathbf{A}$, the commutation matrix $\mathbf{I}_{(p,q)}$ is defined by vec $\mathbf{A}^t = \mathbf{I}_{(p,q)}$vec $\mathbf{A}$, where "vec" transforms the matrix $\mathbf{A}$ into a vector by stacking the columns of $\mathbf{A}$ onto each other.

LEMMA A.1.

$$\text{(i)} \quad \operatorname{vec}(\mathbf{BXC}) = (\mathbf{C}^t \otimes \mathbf{B}) \operatorname{vec} \mathbf{X},$$

$$\text{(ii)} \quad \operatorname{vec}(\mathbf{BC}) = (\mathbf{I} \otimes \mathbf{B}) \operatorname{vec} \mathbf{C} = (\mathbf{C}^t \otimes \mathbf{I}) \operatorname{vec} \mathbf{B} = (\mathbf{C}^t \otimes \mathbf{B}) \operatorname{vec} \mathbf{I},$$

$$\text{(iii)} \quad \operatorname{tr}(\mathbf{BCD}) = (\operatorname{vec} \mathbf{B})^t \ (\mathbf{I} \otimes \mathbf{C}) \operatorname{vec} \mathbf{D},$$

$$\text{(iv)} \quad \operatorname{tr}(\mathbf{BX}^t\mathbf{CXD}) = (\operatorname{vec} \mathbf{X})^t \ (\mathbf{B}^t\mathbf{D}^t \otimes \mathbf{C}) \operatorname{vec} \mathbf{X}$$
$$= (\operatorname{vec} \mathbf{X})^t \ (\mathbf{DB} \otimes \mathbf{C}^t) \operatorname{vec} \mathbf{X}.$$

LEMMA A.2.

$$\text{(i)} \quad \mathbf{I}_{(p,1)} = \mathbf{I}_{(1,p)} = \mathbf{I}_p,$$

$$\text{(ii)} \quad \mathbf{I}^t_{(p,q)} = \mathbf{I}_{(q,p)},$$

$$\text{(iii)} \quad \mathbf{I}_{(p,q)}\mathbf{I}_{(q,p)} = \mathbf{I}_{pq},$$

$$\text{(iv)} \quad \mathbf{B} \otimes \mathbf{A} = \mathbf{I}_{(p,r)}(\mathbf{A} \otimes \mathbf{B})\, \mathbf{I}_{(s,q)} \quad \textit{for} \ \mathbf{A}(p \times q) \quad \textit{and} \quad \mathbf{B}(r \times s).$$

For a symmetric matrix $\mathbf{A}$ the vech-operator does the same as "vec" but uses only the elements a_{ij} for $i \geqslant j$. The relationship between "vec" and "vech" can be described by the elimination matrix $\mathbf{H}$ and the duplication matrix $\mathbf{G}$ as

$$\operatorname{vech} \mathbf{A} = \mathbf{H}_p \operatorname{vec} \mathbf{A} \qquad \text{and} \qquad \operatorname{vec} \mathbf{A} = \mathbf{G}_p \operatorname{vech} \mathbf{A}.$$

Several choices for defining $\mathbf{H}_p$ are possible, depending on how the elements a_{ij} and a_{ji} $(i \neq j)$ are used. For our purpose it was convenient to use $\frac{1}{2}$ as entries in $\mathbf{H}_p$; e.g., for $p = 3$ we then have

$$\mathbf{H}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Notice that this definition of the elimination matrix is not the same as in Magnus (1988), but rather follows Henderson and Searle (1979). The results of Lemma 3.1, Theorem 3.3, and Lemma A.3 depend on this particular choice of $\mathbf{H}_p$.

LEMMA A.3.

(i) $$\mathbf{H}_p \mathbf{G}_p = \mathbf{I}_{p(p+1)/2},$$

(ii) $$\mathbf{I}_{(p,\, p)} \mathbf{G}_p = \mathbf{G}_p,$$

(iii) $$\mathbf{H}_p \mathbf{I}_{(p,\, p)} = \mathbf{H}_p,$$

(iv) $$\mathbf{H}_p = (\mathbf{G}_p^t\, \mathbf{G}_p)^{-1}\, \mathbf{G}_p^t,$$

(v) $$2\mathbf{G}_p \mathbf{H}_p = \mathbf{I}_{p^2} + \mathbf{I}_{(p,\, p)}.$$

The last results relate to derivatives of matrix-valued functions.

LEMMA A.4. *Let* $\mathbf{X}$ *be a symmetric p by p matrix of functionally independent variables (except* $x_{ij} = x_{ji}$*). Then*

(a) $$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1}\, \mathbf{\Delta}_{ij}^*\, \mathbf{X}^{-1},$$

where $\mathbf{\Delta}_{ij}^*$ *is* 1 *in cells* (i, j) *and* (j, i) *and* 0 *otherwise.*

(b) $$\frac{\partial \log \det \mathbf{X}}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \operatorname{diag}(\mathbf{X}^{-1}).$$

(c) *Let* $\mathbf{A}$ *be an arbitrary p by p matrix, then*

$$\frac{\partial\, \operatorname{tr}(\mathbf{XA})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^t - \operatorname{diag} \mathbf{A}.$$

APPENDIX B: DERIVATIVES INVOLVING ORTHOGONAL MATRICES

For an illustration of Theorem 3.5 we show how the matrix

$$\mathbf{J}_{\operatorname{vec} \beta,\, \beta^*} = \frac{\partial \operatorname{vec} \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{*t}}$$

is determined by the orthogonality constraints of $\boldsymbol{\beta}$. The case $p = 3$ is discussed in detail.

Step 1. We choose $p(p-1)/2$ functionally independent elements of $\boldsymbol{\beta}$. These elements define the vector $\boldsymbol{\beta}^*$. The vector $\boldsymbol{\beta}_c^*$ contains the remaining elements. For $p=3$, let

$$\boldsymbol{\beta}=\begin{pmatrix} a & b & d \\ x & c & e \\ y & z & f \end{pmatrix}$$

be the orthogonal matrix and let

$$\boldsymbol{\beta}^*=(x, y, z)' \qquad \text{and} \qquad \boldsymbol{\beta}_c^*=(a, b, c, d, e, f)'.$$

Step 2. For $i=1, ..., p$, $j=i, ..., p$, the $p(p+1)/2$ orthogonality constraints are given by

$$\boldsymbol{\beta}_i^t\, \boldsymbol{\beta}_j=\begin{cases} 1 & i=j \\ 0 & i\neq j. \end{cases}$$

For $p=3$, the six orthogonality constraints are

$$\begin{aligned}
\boldsymbol{\beta}_1^t\, \boldsymbol{\beta}_1 &= a^2+x^2+y^2=1, \\
\boldsymbol{\beta}_1^t\, \boldsymbol{\beta}_2 &= ab+xc+yz=0, \\
\boldsymbol{\beta}_1^t\, \boldsymbol{\beta}_3 &= ad+xe+yf=0, \\
\boldsymbol{\beta}_2^t\, \boldsymbol{\beta}_2 &= b^2+c^2+z^2=1, \\
\boldsymbol{\beta}_2^t\, \boldsymbol{\beta}_3 &= bd+ce+zf=0, \\
\boldsymbol{\beta}_3^t\, \boldsymbol{\beta}_3 &= d^2+e^2+f^2=1.
\end{aligned}$$

Step 3. Taking partial derivatives with respect to an arbitrary scalar v in the orthogonality equations of Step 2 yields

$$\mathbf{A}\cdot\frac{\partial \operatorname{vec} \boldsymbol{\beta}}{\partial v}=\mathbf{0},$$

where the matrix $\mathbf{A}$ has dimension $p(p+1)/2$ by p^2. We note that each row in $\mathbf{A}$ corresponds to a particular orthogonality constraint, and it contains the elements of one or two columns involved in the orthogonality constraint. For $p=3$, we obtain

$$\begin{aligned}
2aa_v+2xx_v+2yy_v&=0, \\
a_vb+ab_v+x_vc+xc_v+y_vz+yz_v&=0, \\
a_vd+ad_v+x_ve+xe_v+y_vf+yf_v&=0,
\end{aligned}$$

$$2bb_v + 2cc_v + 2zz_v = 0,$$
$$b_v d + bd_v + c_v e + ce_v + z_v f + zf_v = 0,$$
$$2dd_v + 2ee_v + 2ff_v = 0,$$

or, in matrix notation,

$$\begin{pmatrix} 2a & 2x & 2y & 0 & 0 & 0 & 0 & 0 & 0 \\ b & c & z & a & x & y & 0 & 0 & 0 \\ d & e & f & 0 & 0 & 0 & a & x & y \\ 0 & 0 & 0 & 2b & 2c & 2z & 0 & 0 & 0 \\ 0 & 0 & 0 & d & e & f & b & c & z \\ 0 & 0 & 0 & 0 & 0 & 0 & 2d & 2e & 2f \end{pmatrix} \begin{pmatrix} a_v \\ x_v \\ y_v \\ b_v \\ c_v \\ z_v \\ d_v \\ e_v \\ f_v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Step 4. Since $\boldsymbol{\beta}^*$ contains functionally independent elements, we have

$$\frac{\partial \boldsymbol{\beta}^*}{\partial \boldsymbol{\beta}^{*t}} = \mathbf{I}_{p(p-1)/2}.$$

Using this, and substituting successively all the elements in $\boldsymbol{\beta}^*$ for v, the equation in Step 3 can be written as

$$\mathbf{A}^* \frac{\partial \boldsymbol{\beta}_c^*}{\partial \boldsymbol{\beta}^{*t}} = -\mathbf{C}.$$

The matrix $\mathbf{A}^*$, $p(p+1)/2$ by $p(p+1)/2$, is the same as $\mathbf{A}$ with those columns deleted that correspond to the elements in $\boldsymbol{\beta}^*$. The matrix $\mathbf{C}$ contains the deleted columns. For $p = 3$,

$$\begin{pmatrix} 2a & 0 & 0 & 0 & 0 & 0 \\ b & a & x & 0 & 0 & 0 \\ d & 0 & 0 & a & x & y \\ 0 & 2b & 2c & 0 & 0 & 0 \\ 0 & d & e & b & c & z \\ 0 & 0 & 0 & 2d & 2e & 2f \end{pmatrix} \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \\ d_x & d_y & d_z \\ e_x & e_y & e_z \\ f_x & f_y & f_z \end{pmatrix} = -\begin{pmatrix} 2x & 2y & 0 \\ c & z & y \\ e & f & 0 \\ 0 & 0 & 2z \\ 0 & 0 & f \\ 0 & 0 & 0 \end{pmatrix}$$

Step 5. Solving the equation in Step 4 yields $\partial\boldsymbol{\beta}_c^*/\partial\boldsymbol{\beta}^{*t} = -\mathbf{A}^{*-1}\mathbf{C}$.

Step 6. The matrix $\partial \operatorname{vec} \boldsymbol{\beta}/\partial\boldsymbol{\beta}^{*t}$ is now given by including the rows of

$$\frac{\partial\boldsymbol{\beta}^*}{\partial\boldsymbol{\beta}^{*t}} = \mathbf{I}_{p(p-1)/2}$$

in the matrix $\partial\boldsymbol{\beta}_c^*/\partial\boldsymbol{\beta}^{*t}$ of Step 5. For $p=3$, we obtain the matrix

$$\mathbf{J}_{\operatorname{vec}\beta,\,\beta^*} = \begin{pmatrix} a_x & a_y & a_z \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ b_x & b_y & b_z \\ c_x & c_y & c_z \\ 0 & 0 & 1 \\ d_x & d_y & d_z \\ e_x & e_y & e_z \\ f_x & f_y & f_z \end{pmatrix}.$$

To find the estimated asymptotic standard deviations of all elements of $\hat{\boldsymbol{\beta}}$ with the help of Theorem 3.5, we need to choose different versions of $\boldsymbol{\beta}^*$. For $p \geqslant 3$, three versions are sufficient. One possibility is to choose the lower triangle, the upper triangle, and the diagonal (plus $p(p-3)/2$ other elements) of $\hat{\boldsymbol{\beta}}$. For $p=2$, it suffices to choose two different elements of the orthogonal matrix $\hat{\boldsymbol{\beta}}$.

7. PROOFS

Proof of Lemma 2.2. We introduce Lagrange multipliers δ_h and δ_{hg} $(h, g = 1, ..., p)$ for the constraints in $\boldsymbol{\beta}$. Using

$$\operatorname{tr}(\mathbf{C}_j\boldsymbol{\beta}^t\mathbf{A}_j\boldsymbol{\beta}) = \sum_{h=1}^{p} c_{jh}\boldsymbol{\beta}_h^t\mathbf{A}_j\boldsymbol{\beta}_h,$$

and setting the first derivative with respect to $\boldsymbol{\beta}_l$ equal to zero implies

$$\sum_{j=1}^{J} c_{jl}(\mathbf{A}_j + \mathbf{A}_j^t)\,\boldsymbol{\beta}_l - 2\delta_l\boldsymbol{\beta}_l - 2\sum_{h\neq l}\delta_{hl}\boldsymbol{\beta}_h = \mathbf{0},$$

where we use $\delta_{lh} = \delta_{hl}$. Multiplying from the left by $\boldsymbol{\beta}_m^t$ $(m \neq l)$ yields

$$\sum_{j=1}^{J} c_{jl} \boldsymbol{\beta}_m^t (\mathbf{A}_j + \mathbf{A}_j^t)\, \boldsymbol{\beta}_l = 2\delta_{ml},$$

and similarly

$$\sum_{j=1}^{J} c_{jm} \boldsymbol{\beta}_l^t (\mathbf{A}_j + \mathbf{A}_j^t)\, \boldsymbol{\beta}_m = 2\delta_{lm}.$$

Equating the last two expressions gives

$$\boldsymbol{\beta}_m^t \left[\sum_{j=1}^{J} (c_{jl} - c_{jm})(\mathbf{A}_j + \mathbf{A}_j^t) \right] \boldsymbol{\beta}_l = 0.$$

Proof of Lemma 3.1. (i) Recall that for a symmetric p by p matrix $\mathbf{A}$, vech $\mathbf{A} = \mathbf{H}_p$ vec $\mathbf{A}$ and vec $\mathbf{A} = \mathbf{G}_p$ vech $\mathbf{A}$. Thus the $p(p+1)/2$ by p^2 matrix $\mathbf{H}_p$ is defined as follows: the elements in row $(h-1)\, p - h(h-1)/2 + g$ and columns $(h-1)\, p + g$ and $(g-1)\, p + h$ are 1 for $h = g$ and $\frac{1}{2}$ for $h \neq g$ $(h = 1, ..., p; g = h, ..., p)$, and the other elements are zero. The case is similar for the p^2 by $p(p+1)/2$ matrix $\mathbf{G}_p$: the elements in column $(h-1)$ $p - h(h-1)/2 + g$ and rows $(h-1)\, p + g$ and $(g-1)\, p + h$ are 1 $(h = 1, ..., p;$ $g = h, ..., p)$, and the other elements are 0.

The matrix $2\mathbf{I}_{p^2} - \mathbf{D}_p$ is diagonal with diagonal elements equal to 2 except for positions $(h-1)\, p + h$ $(h = 1, ..., p)$, where the elements equal 1. Multiplying $\mathbf{H}_p$ from the right by $2\mathbf{I}_{p^2} - \mathbf{D}_p$ results in multiplying each column of $\mathbf{H}_p$ by the corresponding element of $2\mathbf{I}_{p^2} - \mathbf{D}_p$ which is 2 or 1. But by the structure of $\mathbf{H}_p$ this means that each element $\frac{1}{2}$ is multiplied by 2 whereas each element 1 is multiplied by 1. It follows that the product $\mathbf{H}_p(2\mathbf{I}_{p^2} - \mathbf{D}_p)$ has the same structure as $\mathbf{H}_p$ except that $\frac{1}{2}$ is replaced by 1. By definition this is exactly the matrix $\mathbf{G}_p^t$.

(ii) This is an immediate consequence of (i) and Lemma A.3(iii).

(iii) By Lemma A.2(iii, iv) and Lemma A.3(ii) we have

$$\begin{aligned}
\mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A})(\mathbf{I}_{p^2} + \mathbf{I}_{(p,\,p)}) &= \mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A}) + \mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A})\, \mathbf{I}_{(p,\,p)} \\
&= \mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A}) + \mathbf{G}_p^t \mathbf{I}_{(p,\,p)}(\mathbf{A} \otimes \mathbf{A}) \\
&= \mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A}) + \mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A}) \\
&= 2\mathbf{G}_p^t(\mathbf{A} \otimes \mathbf{A}).
\end{aligned}$$

(iv) This is the same proof as that in (iii).

Proof of Lemma 3.2. (i) By Lemma A.4 we have $\partial \mathbf{X}^{-1}/\partial x_{ij} = -\mathbf{X}^{-1}\,\mathbf{\Delta}_{ij}^{*}\,\mathbf{X}^{-1}$, which, by Lemma A.1, implies

$$\frac{\partial \operatorname{vec} \mathbf{X}^{-1}}{\partial x_{ij}} = -\operatorname{vec}(\mathbf{X}^{-1}\,\mathbf{\Delta}_{ij}^{*}\,\mathbf{X}^{-1}) = -(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}) \operatorname{vec} \mathbf{\Delta}_{ij}^{*},$$

and

$$\frac{\partial \operatorname{vec} \mathbf{X}^{-1}}{\partial (\operatorname{vec} \mathbf{X})^{t}} = -(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})(\operatorname{vec} \mathbf{\Delta}_{11}^{*}, \operatorname{vec} \mathbf{\Delta}_{21}^{*}, ..., \operatorname{vec} \mathbf{\Delta}_{pp}^{*}).$$

Since the elements of $\mathbf{\Delta}_{ij}^{*}$ are zero except for positions (i, j) and (j, i), we see that

$$\begin{aligned}
&(\operatorname{vec} \mathbf{\Delta}_{11}^{*}, \operatorname{vec} \mathbf{\Delta}_{21}^{*}, ..., \operatorname{vec} \mathbf{\Delta}_{pp}^{*}) \\
&\quad = (\operatorname{vec} \mathbf{\Delta}_{11}, \operatorname{vec} \mathbf{\Delta}_{21}, ..., \operatorname{vec} \mathbf{\Delta}_{pp}) + (\operatorname{vec} \mathbf{\Delta}_{11}^{t}, \operatorname{vec} \mathbf{\Delta}_{21}^{t}, ..., \operatorname{vec} \mathbf{\Delta}_{pp}^{t}) - \mathbf{D}_{p},
\end{aligned}$$

where we use $\mathbf{\Delta}_{ij}$ for the matrix with 1 in position (i, j), and 0 elsewhere. By

$$(\operatorname{vec} \mathbf{\Delta}_{11}, \operatorname{vec} \mathbf{\Delta}_{21}, ..., \operatorname{vec} \mathbf{\Delta}_{pp}) = \mathbf{I}_{p^2},$$

and $\operatorname{vec} \mathbf{\Delta}_{ij}^{t} = \mathbf{I}_{(p,\,p)} \operatorname{vec} \mathbf{\Delta}_{ij}$, we obtain

$$\frac{\partial \operatorname{vec} \mathbf{X}^{-1}}{\partial (\operatorname{vec} \mathbf{X})^{t}} = -(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})(\mathbf{I}_{p^2} + \mathbf{I}_{(p,\,p)} - \mathbf{D}_{p}).$$

(ii) This follows by (i) and Lemma 3.1(ii):

$$\frac{\partial \operatorname{vec} \mathbf{X}^{-1}}{\partial (\operatorname{vec} \mathbf{X})^{t}} = -(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})(\mathbf{I}_{p^2} + \mathbf{I}_{(p,\,p)} - \mathbf{D}_{p}),$$

which implies

$$\begin{aligned}
\frac{\partial \operatorname{vech} \mathbf{X}^{-1}}{\partial (\operatorname{vech} \mathbf{X})^{t}} &= -\mathbf{H}_{p}\,(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})(\mathbf{I}_{p^2} + \mathbf{I}_{(p,\,p)} - \mathbf{D}_{p})\,\mathbf{H}_{p}^{t} \\
&= -\mathbf{H}_{p}\,(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})\,\mathbf{G}_{p}.
\end{aligned}$$

Proof of Lemma 3.4. The structure of $\mathbf{U}$ implies

$$\mathbf{b}^{t}\mathbf{U}\mathbf{b} = \sum_{h=1}^{p} u_{h} \boldsymbol{\beta}_{h}^{t}\, \boldsymbol{\beta}_{h} = \sum_{h=1}^{p} u_{h},$$

which does not depend on $\boldsymbol{\beta}^*$. Consequently $\partial \mathbf{b}^t \mathbf{U} \mathbf{b} / \partial \boldsymbol{\beta}^* = \mathbf{0}$. On the other hand, we have

$$\frac{\partial \mathbf{b}^t \mathbf{U} \mathbf{b}}{\partial \boldsymbol{\beta}^*} = 2 \mathbf{J}^t_{b,\, \beta^*} \mathbf{U} \mathbf{b} = 2 \frac{\partial \mathbf{b}^t}{\partial \boldsymbol{\beta}^*} \mathbf{U} \mathbf{b},$$

from which it follows that $\mathbf{c}^t_{ml} \mathbf{U} \mathbf{b} = 0$. By taking derivatives again we then obtain

$$\frac{\partial \mathbf{c}^t_{ml} \mathbf{U} \mathbf{b}}{\partial \boldsymbol{\beta}^*} = \mathbf{0},$$

which implies $\mathbf{J}^t_{c_{ml},\, \beta^*} \mathbf{U} \mathbf{b} + \mathbf{J}^t_{b,\, \beta^*} \mathbf{U} \mathbf{c}_{ml} = \mathbf{0}$. This is equivalent to

$$\frac{\partial \mathbf{c}^t_{ml}}{\partial \boldsymbol{\beta}^*} \mathbf{U} \mathbf{b} = -\frac{\partial \mathbf{b}^t}{\partial \boldsymbol{\beta}^*} \mathbf{U} \mathbf{c}_{ml}.$$

ACKNOWLEDGMENTS

The work of B.E.N. was supported in part by a postdoctoral grant from the Swiss National Science Foundation. B.D.F. was supported in part by NSF Grant DMS-9802522. The authors thank two referees for their helpful comments on an earlier version of this article. In particular, one referee pointed out some possible confusion about different definitions of the elimination matrix.

REFERENCES

1. T. W. Anderson, Asymptotic theory for principal component analysis, *Ann. Math. Statist.* **34** (1963), 122–148.
2. D. J. Donnell, A. Buja, and W. Stuetzle, Analysis of additive dependencies and concurvities using smallest additive principal components, *Ann. Statist.* **22** (1994), 1635–1673.
3. B. Flury, "Common Principal Components and Related Multivariate Models," Wiley, New York, 1988.
4. B. D. Flury and B. E. Neuenschwander, Principal component models for patterned covariance matrices, with applications to canonical correlation analysis of several sets of variables, *in* "Recent Advances in Descriptive Multivariate Analysis" (W. J. Krzanowski, Ed.), pp. 90–112, Oxford Science Publications, Oxford, UK, 1995a.
5. B. D. Flury and B. E. Neuenschwander, Simultaneous diagonalization algorithms with applications in multivariate statistics, *in* "Approximation and Computation" (R. V. M. Zahar, Ed.), pp. 179–205, Basel, Birkhäuser, 1995b.
6. B. Flury and M. J. Schmid, Quadratic discriminant functions with constraints on the covariance matrices: Some asymptotic results, *J. Multivariate Anal.* **40** (1992), 244–261.
7. R. Gnanadesikan, "Methods for Statistical Data Analysis of Multivariate Observations," Wiley, New York, 1977.

8. R. Gnanadesikan and M. B. Wilk, Data analytic methods in multivariate statistical analysis, *in* "Multivariate Analysis" (P. R. Krishnaiah, Ed.), Vol. II, pp. 593–638, Academic Press, New York, 1969.
9. I. J. Good, Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Ann. Math. Statist.* **34** (1963), 911–934.
10. R. F. Gunst and R. L. Mason, Biased estimation in regression: an evaluation using mean squared error, *J. Amer. Statist. Assoc.* **72** (1977), 616–628.
11. T. Hastie and W. Stuetzle, Principal curves, *J. Amer. Statist. Assoc.* **84** (1989), 502–516.
12. H. V. Henderson and S. R. Searle, Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics, *Canad. J. Statist.* **7** (1979), 65–81.
13. H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educational Psych.* **24** (1933), 417–441.
14. C. P. Klingenberg, B. E. Neuenschwander, and B. D. Flury, Ontogeny and individual variation: Analysis of patterned covariance matrices with common principal components, *Systematic Biol.* **45**, No. 2 (1996), 135–150.
15. W. J. Krzanowski, Between-groups comparision of principal components, *J. Amer. Statist. Assoc.* **74** (1979), 703–707.
16. S. Kullback, "Information Theory and Statistics," Wiley, New York, 1959.
17. J. R. Magnus, "Linear Structures," Charles Griffin, London, 1988.
18. E. R. Mansfield, J. T. Webster, and R. F. Gunst, An analytic variable selection procedure for principal component regression, *App. Statist.* **26** (1977), 34–40.
19. B. E. Neuenschwander, "Common Principal Components for Dependent Random Vectors," Ph.D. thesis, Department of Mathematical Statistics, University of Berne, Switzerland.
20. B. E. Neuenschwander and B. D. Flury, Common canonical variates, *Biometrika* **82** (1995), 553–560.
21. K. Pearson, On lines and planes of closest fit to systems of points in space, *Philos. Magazine Ser. 6* **2** (1901), 559–572.
22. R. J. Serfling, "Approximation Theorems of Mathematical Statistics," Wiley, New York, 1980.
23. T. Tarpey, L. Li, and B. D. Flury, Principal points and self-consistent points of elliptical distributions, *Ann Statist.* **23** (1995), 103–112.
24. R. Tibshirani, Principal curves revisited, *Statist and Comput.* **2** (1992), 183–190.